

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Big Data

Lección 3

PyCaret y AutoML

ÍNDICE

- ✓ Introducción
- ✓ Objetivos
- ✓ PyCaret te automatiza muchas cosas en una sola línea
- ✓ PyCaret te aporta mucha información
- ✓ Kaggle: Nueva predicción para el Titanic DataSet

INTRODUCCIÓN

En esta lección haremos una breve introducción a PyCaret y al AutoML

OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Trabajar con PyCaret para Clasificación binaria
- 2 Conocer más algoritmos de Clasificación
- 3 Entender de la importancia del AutoML en Data Science

PyCaret te automatiza muchas cosas en una sola línea

step 4 - compare_models()

```
In [7]: # k-Fold cross validation
# este concepto mejor esperar a Machine Learning por la explicación técnica
compare_models()
```

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
gbc	Gradient Boosting Classifier	0.8412	0.8831	0.6966	0.8460	0.7601	0.6436	0.6534	0.2680
lightgbm	Light Gradient Boosting Machine	0.8332	0.8810	0.7190	0.8140	0.7560	0.6310	0.6396	0.3110
rf	Random Forest Classifier	0.8315	0.8830	0.7666	0.7804	0.7674	0.6358	0.6416	0.4280
lr	Logistic Regression	0.8250	0.8723	0.7271	0.7876	0.7510	0.6169	0.6225	1.8440
et	Extra Trees Classifier	0.8186	0.8622	0.7318	0.7769	0.7480	0.6070	0.6132	0.3830
ridge	Ridge Classifier	0.8106	0.0000	0.7134	0.7622	0.7330	0.5867	0.5910	0.0230
lda	Linear Discriminant Analysis	0.8074	0.8598	0.6962	0.7686	0.7249	0.5776	0.5843	0.0580
ada	Ada Boost Classifier	0.8042	0.8524	0.7186	0.7467	0.7263	0.5746	0.5802	0.2020
dt	Decision Tree Classifier	0.7946	0.7744	0.7184	0.7191	0.7155	0.5553	0.5580	0.0310
knn	K Neighbors Classifier	0.7047	0.7318	0.5059	0.6075	0.5488	0.3339	0.3381	0.0750
svm	SVM - Linear Kernel	0.6875	0.0000	0.7180	0.6163	0.6238	0.3768	0.4152	0.0310
nb	Naive Bayes	0.6629	0.7913	0.1152	0.7583	0.1977	0.1128	0.1940	0.0270
qda	Quadratic Discriminant Analysis	0.6339	0.5963	0.4561	0.5398	0.4522	0.1996	0.2010	0.0470

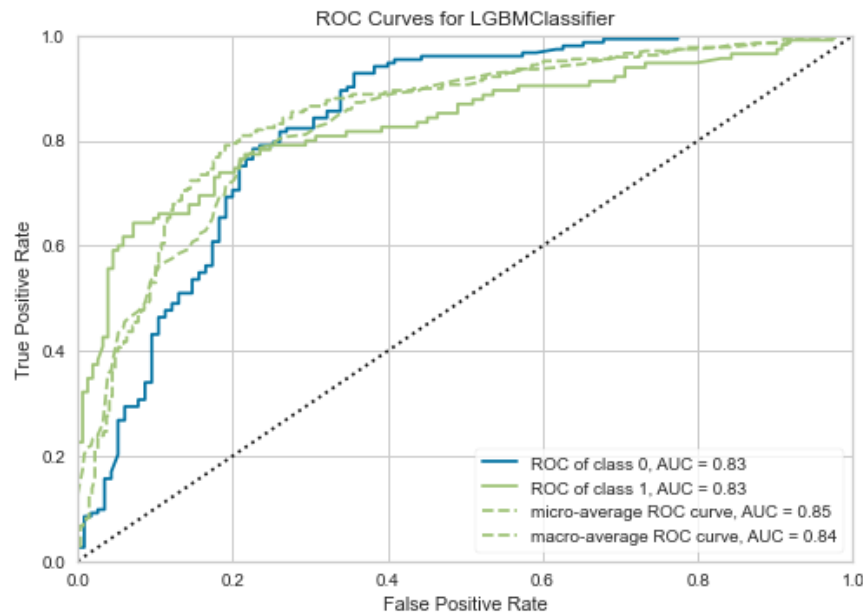
```
Out[7]: GradientBoostingClassifier(ccp_alpha=0.0, criterion='friedman_mse', init=None,
learning_rate=0.1, loss='deviance', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
```

PyCaret te aporta mucha información

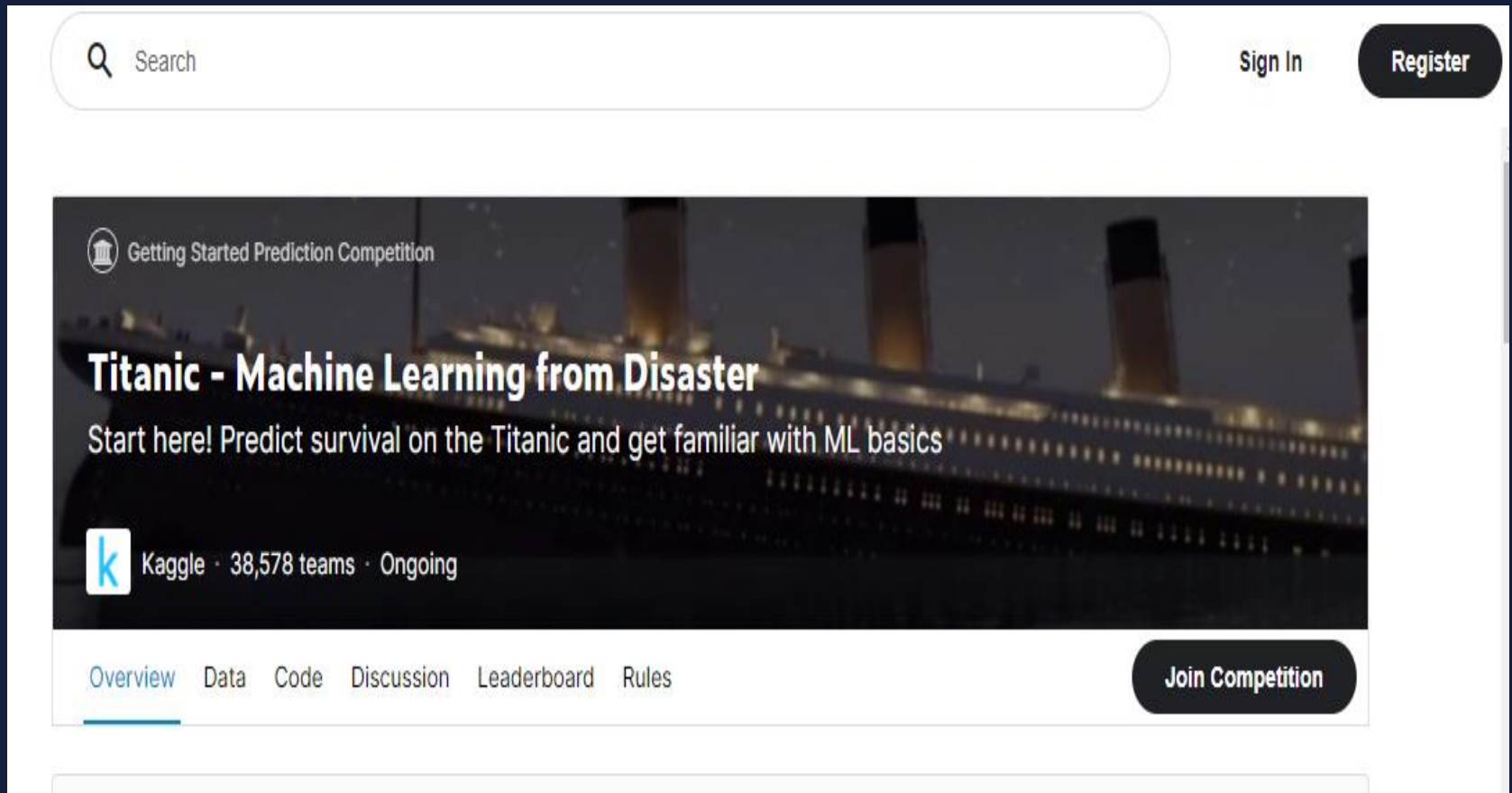
step 6 - evaluate_model() para el mejor

```
In [21]: evaluate_model(tune_lightgbm_acc)
```

Plot Type:	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall
	Prediction Error	Class Report	Feature Selection	Learning Curve	Manifold Learning
	Calibration Curve	Validation Curve	Dimensions	Feature Importance	Feature Importance...
	Decision Boundary	Lift Chart	Gain Chart	Decision Tree	



Kaggle: Nueva predicción para el Titanic DataSet



The image shows the Kaggle competition page for "Titanic - Machine Learning from Disaster". At the top, there is a search bar with a magnifying glass icon and the word "Search", and two buttons: "Sign In" and "Register". Below the search bar is a large banner featuring a night-time photograph of the Titanic ship. The banner contains the text "Getting Started Prediction Competition" with a small icon, the title "Titanic - Machine Learning from Disaster" in large white font, and the subtitle "Start here! Predict survival on the Titanic and get familiar with ML basics". Below the banner, the Kaggle logo is followed by the text "Kaggle · 38,578 teams · Ongoing". At the bottom of the banner area, there are links for "Overview", "Data", "Code", "Discussion", "Leaderboard", and "Rules", with "Overview" being the active link. A "Join Competition" button is located in the bottom right corner of the banner area.

MUCHAS GRACIAS POR SU ATENCIÓN



jmpena@grupomainjobs.com



José Manuel Peña

<https://www.linkedin.com/in/jos%C3%A9-manuel-pe%C3%B1a-castro-7566b349/>



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados